

METADATI E CONSERVAZIONE DIGITALE

Roma, 3 aprile 2007

Maria Guercio, Università di Urbino

“IL MONDO VOLATILE DEI METADATI”

- Gli sviluppi degli ultimi anni sono caratterizzati da una proliferazione e continua evoluzione delle strutture (formati) dei metadati per il trattamento delle risorse digitali (documenti, oggetti) e per sostenere i successivi processi conservativi
- La letteratura in materia ha cercato (in modo finora insoddisfacente, ripetitivo e inutilmente sovrabbondante) di definirne i confini, la natura, le funzioni

I METADATI PER LA CONSERVAZIONE SECONDO LO STANDARD NISO

- Di particolare rilevanza (ma non di altrettanta efficacia) lo sforzo fatto in questi anni nel campo dei *metadati per la conservazione digitale* da parte di alcune comunità di pratiche (dai ricercatori in campo scientifico in prima istanza e poi dai bibliotecari)
- Gli sforzi congiunti hanno condotto a un'ipotesi di classificazione non particolarmente felice ma ormai largamente utilizzata e tradotta nelle norme NISO 2004 anche a seguito della sua adozione da parte dello standard OAIS.
- Di fatto negli sviluppi implementativi i metadati per la conservazione in quanto “informazioni necessarie per archiviare e conservare una risorsa al fine di assicurarne l'autenticità e la possibilità di riproduzione/ricostituzione” si limitano a identificare e gestire informazioni di natura quasi esclusivamente tecnologica e comunque difficilmente riferibili ad archivi digitali complessi.

I METADATI PER LA CONSERVAZIONE: nella definizione

- “Information that supports and documents the process of digital preservation: the term is usually reserved for metadata that specifically supports the *functions* of maintaining the *fixity*, *viability*, *renderability*, *understandability*, and/or *authenticity* of a digital material *in a preservation context*”

(da Priscilla Caplan, *Preservation metadata. Report for DCC*, 2006)

I METADATI PER LA CONSERVAZIONE: nelle implementazioni

- Che cosa hanno finito per includere:
 - elementi di metadati amministrativi
 - elementi di metadati strutturali tecnici (parte di quelli amministrativi) che documentano caratteristiche dettagliate dei formati e dell'ambiente d'uso)
 - eventualmente anche la documentazione relativa alla gestione dei diritti (proprietà intellettuale, accesso, ecc.)

(cfr L. Brian, R. Gartner, *Technology Watch Report: Preservation Metadata*, sept. 2005,
<http://www.dpconline.org/docs/reports/dpctw05-01.pdf>)

I METADATI PER LA CONSERVAZIONE: UNA RICOSTRUZIONE STORICA, 1996-2006

Il rapporto *Preserving Digital Information* del 1996

- 1996: pubblicazione del rapporto *Preserving Digital Information*
Il fine del processo conservativo è *l'integrità dell'oggetto*. Gli *attributi* che la garantiscono sono espressi in forme molto generali:
 - *Contenuto* inteso non solo come flusso di bit (necessariamente modificabile per risolvere i nodi dell'obsolescenza) ma anche come struttura logica cui ricondurre il contenuto stesso
 - *Fixity*: meccanismi per prevenire o scoprire modifiche non desiderate nel deposito digitale (ad esempio firma digitale, checksum)
 - *Reference*: meccanismi per identificare e localizzare le risorse (nel deposito)
 - *Provenance*: meccanismi che memorizzano la catena della custodia
 - *Context*: si riferisce ai modi in cui le risorse digitali interagiscono con altre risorse, hardware, software e supporti inclusi

I METADATI PER LA CONSERVAZIONE: UNA RICOSTRUZIONE STORICA, 1996-2006 1999-2003: lo standard OAIS

- Il modello OAIS, approvato come standard ISO 14721, ha trasformato gli attributi in un modello concettuale funzionale e informativo coerente
- Il modello ha notevoli potenzialità di sviluppo anche se le applicazioni finora realizzate hanno riguardato collezioni di limitata complessità e non si sono tradotti in processi automatici di acquisizione
- 2002-2003, Preservation Metadata Framework: i set di metadati sviluppati in conformità con lo standard ISO risentono della loro origine biblioteconomica (METS, MAG)

I METADATI PER LA CONSERVAZIONE: UNA RICOSTRUZIONE STORICA, 1996-2006 2003-2006 PREMIS

- E' sviluppato coerentemente con le precedenti attività di normalizzazione con il compito pragmatico di:
 - sviluppare un set di elementi *cruciali e facilmente implementabili* per la conservazione di oggetti digitali e, in generale, per la gestione di depositi digitali
- I primi risultati sono stati:
 - un insieme essenziale di componenti informative coerente con set di metadati descrittivi esistenti in domini specifici (METS, DUBLIN CORE, EAD, ecc.)
 - un data dictionary finalizzato a facilitare l'uso dello schema elaborato

I METADATI PER LA CONSERVAZIONE: UNA RICOSTRUZIONE STORICA, 1996-2006 2003-2006 GLI ELEMENTI DI PREMIS

Il modello di dati individua cinque tipi di entità rilevanti nelle attività conservative, riconducibili (ma non sovrapponibili) alle categorie OAIS :

- Intellectual entity
- Object: file, bitstream, rappresentazione
- Event
- Right
- Agent

I METADATI PER LA CONSERVAZIONE: UNA RICOSTRUZIONE STORICA, 1996-2006 2003-2006 LE CRITICITA' DI PREMIS

- Premis si limita a descrivere gli oggetti digitali e gli eventi che interagiscono con gli oggetti medesimi nei processi conservativi: le componenti descrittive e l'analisi degli agenti sono affidate agli esperti di ciascun dominio e agli standard di settore (ad esempio ISAD/ISAAR per la materia archivistica)
- Eccesso di granularità e rischio di frammentarietà: gli stessi partecipanti al gruppo di lavoro riconoscono la complessità del modello utilizzato, cosiddetto a “cipolla”, ovvero un sistema di raccolta, conservazione, gestione e aggiornamento dei metadati necessari per strati successivi collegati in sequenza
- Per il patrimonio archivistico risultano insufficienti i meccanismi di correlazione e di gestione delle relazioni gerarchiche
- Da sottolineare che il termine *Representation* utilizzato in OAIS in quanto *Representation Information* assume in Premis una connotazione esclusivamente tecnica: un insieme di file che include anche i metadati che identificano la struttura, necessari per una completa e ragionevole restituzione dell'entità intellettuale :

I METADATI PER LA CONSERVAZIONE: UNA
RICOSTRUZIONE STORICA, 1996-2006
2003-2006 L'IMPLEMENTAZIONE DI PREMIS

- Training
- Editorial board group
- Gruppo allargato di implementatori (PIG): gli istituti archivistici sono quasi del tutto assenti
- Gruppo di discussione e wiki per la pubblicazione di esempi
- Schema XML per lo scambio di metadati

<http://www.loc.gov/standards/premis>

I METADATI PER LA CONSERVAZIONE

1996-2006: IL PERCORSO DEGLI ARCHIVISTI

- E' utile osservare che in campo archivistico:
 - l'individuazione analitica del contesto tecnologico e delle sue componenti, così come l'analisi tecnica degli oggetti conservati non è sempre necessariamente rilevante (se non:
 - come *documentazione del processo conservativo* al fine di assicurare *la presunzione di veridicità* della fonte, oppure
 - al fine di amministrare un deposito digitale
 - Per gli archivisti (vedi le conclusioni del progetto InterPares) la conservazione di un documento digitale non implica necessariamente la conservazione di tutte le sue proprietà originali.
 - Una parte considerevole delle le informazioni definite “metadati di conservazione” sono più correttamente trattate in termini di attività descrittive ed elaborate con strumenti di analisi interni al processo di acquisizione, ad esempio nell'ambito della storia della custodia secondo lo standard ISAD
 - Di conseguenza l'utilità di PREMIS (o MAG) è limitata alle esigenze del singolo deposito e può ridursi a un numero molto limitato di elementi

I METADATI PER LA CONSERVAZIONE

1996-2006: IL PERCORSO DEGLI ARCHIVISTI

I PROGETTI DI RICERCA

- Tutti i progetti di ricerca (Pittsburgh, Interpares, VERS, l'attività dell'ISO TC 46, SC 11 sul record management) hanno identificato i requisiti funzionali dei sistemi di produzione dei documenti come la condizione centrale per assicurare la conservazione degli archivi digitali
- I metadati per la conservazione sono quindi inestricabilmente legati ai metadati per la gestione dei documenti, presenti nella fase attiva
- E' ritenuto inutile e sconsigliato elaborare un set universale di metadati (del resto anche PREMIS si limita a individuare "things that *most* working preservation repositories are *likely* to need to know in order to support digital preservation" (anomalie dei formati e presenza di cifrature, ambienti d'uso, strumenti di validazione)

DAL RAPPORTO DI INTERPARES SUI METADATI PER LA CONSERVAZIONE DIGITALE (2006)

- Il processo di acquisizione dei metadati è destinato ad avere peso e successo a fini conservativi se gestito automaticamente, strettamente correlato ai processi di formazione dei documenti e alla gestione delle informazioni di contesto, utilizzato come componente specifica di una catena della conservazione
- I metadati resi disponibili nei processi di formazione dei documenti e rilevanti a fini conservativi sono finalizzati ad acquisire informazioni significative in relazione agli aspetti gestionali e alle transazioni che hanno per oggetto i documenti archivistici, la cui principale funzione è quella di *testimonianza di attività giuridicamente rilevanti* e non di *contenitore di informazioni*
- Inoltre i metadati stessi di per sé non sono sempre (almeno per ora e a lungo) in grado di fornire una visione complessiva e autonoma dei contesti originali di produzione dei documenti e richiedono una mediazione “descrittiva” da parte dell’archivista.
- Da *Investigating the roles and requirements, manifestations and management of metadata in the creation of reliable and preservation of authentic electronic entities*, Report on the work and findings of the Interpares 2 Description Cross Domain Group, 2006

I METADATI NELLA *CHAIN OF PRESERVATION* DI INTERPARES (2006)

- Record Creation Metadata:
 - Identity Metadata
 - Integrity Metadata
 - Creation Process Metadata
- Recordkeeping Metadata
 - Identity Metadata
 - Integrity Metadata
 - Creation Process Metadata
- Record Preservation Metadata
- Transfer Authorization Metadata
 - Transfer Verification Metadata
 - Feasibility Confirmation Metadata
 - Record Accession Metadata
 - Preservation Action Metadata
 - Description Metadata
 - Retrieval Metadata
 - Presentation Metadata
 - Package Output Metadata

METADATI CHE ASSICURANO L'IDENTITÀ DEL DOCUMENTO ARCHIVISTICO (INTERPARES E MOREQ)

1. Nomi delle persone che concorrono alla formazione del documento: autore, scrittore, responsabile dell'indirizzo elettronico e destinatario
2. Oggetto: sintesi del contenuto giuridico del documento
3. Date di produzione e trasmissione: data del documento, data di ricevimento o spedizione (data archivistica), data di trasmissione
4. Espressione del vincolo archivistico (classificazione/fascicolazione)
5. Indicazione di allegati

METADATI CHE DEFINISCONO L'INTEGRITÀ DEL DOCUMENTO ARCHIVISTICO (INTERPARES E MOREQ)

1. Nome dell'ufficio competente a trattare la materia (metadati di assegnazione)
2. Nome dell'ufficio di responsabilità principale per la tenuta del documento
3. Indicazione delle annotazioni aggiunte al documento (metadati che identificano procedure amministrative, ecc.)
4. Indicazione di modifiche tecniche

METADATI PER L'ACCERTAMENTO DELL'IDENTITÀ E DELL'INTEGRITÀ DEL DOCUMENTO ARCHIVISTICO (INTERPARES E MOREQ)

1. *audit trail* o altre procedure protettive che prevengano, scoprano e correggano la perdita o la corruzione dei documenti
2. controllo/monitoraggio relativo al deterioramento dei supporti e ai cambiamenti tecnologici (obsolescenza delle tecnologie)
3. autenticazione: regole specifiche concernenti quali documenti debbano essere autenticati e da chi e i sistemi di autenticazione
4. esistenza di privilegi di accesso concernenti la produzione, la modificazione, l'annotazione, il movimento e la distruzione dei documenti
5. identificazione delle procedure di trasferimento: indicazione dei documenti trasferiti, delle strutture e persone responsabili per il trasferimento e per la sua acquisizione, data del trasferimento, identificativo del versamento, relazioni archivistiche con gli altri documenti e archivi collegati, documentazione di autorizzazione

LA CONCLUSIONE DEL PROGETTO INTERPARES 2: MADRAS

- The Metadata and Archival Description Registry and Analysis System (MADRAS):
 - schema-level (i.e., not a comprehensive element-level) metadata registry, to include an analytical assessment tool that could be used by the researchers to evaluate the current capabilities of registered metadata schemas.

LE FINALITA' DI MADRAS

- Sostenere la registrazione non ambigua di schemi e di set di metadati nonché di profili applicativi rilevanti per la conservazione di entità digitali autentiche
- consentire l'analisi dei materiali registrati rispetto ai requisiti definiti in Interpares e negli standard ISO 23081 Records Management Metadata Standard,
- Sviluppare raccomandazioni e fornire una infrastruttura normalizzata di riferimento per diverse aree applicative e di pratiche
- Fornire un sistema analitico di informazione per il gruppo ISO
- TC46/SC11-WG1 in relazione alla definizione di uno standard sulla conservazione

LO STANDARD ISO 23081

LINEE GUIDA PER I METADATI NELL'AMBITO DEL RM

LO STANDARD ISO 23081.
I METADATI PER DOCUMENTI ARCHIVISTICI

- Definisce con riferimento allo standard ISO 15489 sul record management le caratteristiche generali di un quadro di riferimento per lo sviluppo e l'applicazione di un sistema di metadati per la gestione dei documenti informatici

LO STANDARD ISO 23081. STRUTTURE DI METADATI

Lo standard stabilisce l'obbligo di *predisporre e conservare strutture di metadati* nella forma di veri e propri *schemi* in grado di documentare le relazioni tra i singoli elementi e renderle comprensibili e ricostruibili, finalizzati inoltre a sostenere esigenze di integrazione, interoperabilità, uniformità semantiche e sintattiche, modularità dei sistemi informativi. A titolo esemplificativo lo standard elenca tra gli schemi di entità da sviluppare e conservare:

- schemi di classificazione dei documenti correlati alle funzioni amministrative (piani di classificazione)
- schemi o DTD XML per la definizione di tipologie documentarie, strutture di database e altri oggetti
- schemi logici di database
- schemi per i privilegi di accesso e la sicurezza

LO STANDARD ISO 23081.

TIPOLOGIE DI METADATI

Si prevedono 6 diverse tipologie rispettivamente relative:

- ai documenti: estremi cronologici, persone, struttura dei documenti e degli archivi, formato, caratteristiche tecniche, relazioni interne tra gli elementi costitutivi, requisiti di presentazione/visualizzazione, informazioni sulle condizioni di migrazione o emulazione, informazioni per la gestione e protezione dei supporti, relazioni con i processi amministrativi, informazioni sulle aggregazioni documentarie e/o archivistiche e informazioni sui processi di memorizzazione e archiviazione, sulla sicurezza, sull'accesso, sui sistemi di recupero (ISAD);
- alle regole e policy che governano l'ordinamento giuridico e amministrativo: regolamenti di organizzazione amministrativa, regolamenti relativi al sistema documentario, regolamenti sull'accesso e sulla sicurezza (ISAAR);
- ai soggetti produttori: denominazioni, quadro delle strutture interne, responsabilità e profili utenti per le diverse attività di gestione documentaria (ISAAR);
- ai processi di lavoro (*business process*): funzionigramma e organigramma, quadro dei processi amministrativi in relazione all'iter e alle responsabilità (ISAAR);
- ai processi di gestione documentaria: sistemi di registrazione, piani di classificazione, piani di conservazione, classificazione degli accessi, profili di utenti per la gestione documentaria, piano della sicurezza, repertori dei fascicoli e inventari, documentazione dei processi conservativi (ISAD).
- al documento sui metadati medesimi

LO STANDARD ISO 23081. REGOLAMENTAZIONE SPECIFICA

La regolamentazione in questo campo dovrebbe includere:

- la identificazione delle responsabilità e dei modi per gestirne il coordinamento,
- le relazioni con il complessivo quadro normativo e regolamentare che governa il sistema informativo di un ente con specifico riferimento alla gestione dei metadati,
- l'elenco degli standard adottati,
- le modalità di applicazione degli strumenti per la formazione e l'aggiornamento dei metadati (ad esempio del piano di classificazione e fascicolazione o della produzione di indici),
- gli strumenti di monitoraggio e controllo
- le forme di aggiornamento e i meccanismi per la continuità del sistema.

LO STANDARD ISO 23081. GLI ASPETTI APPLICATIVI

Dal punto di vista applicativo si sottolinea l'esigenza di:

- garantire la sostenibilità di un sistema di metadati in relazione alle continue trasformazioni amministrative e organizzative,
- dedicare particolare attenzione ai problemi della conservazione digitale e alle strategie specifiche necessarie a catturare e mantenere i metadati necessari che comunque devono rispondere a requisiti di normalizzazione e qualità,
- incorporare i metadati di gestione dei documenti nei sistemi archivistici, ovvero nei sistemi finalizzati alla permanenza delle informazioni documentarie.

LO STANDARD ISO 23081.

LE PROCEDURE

Le linee guida dettano procedure per:

- condividere metadati prodotti in altri domini che tengano conto sia delle esigenze della struttura che accoglie il sistema di metadati, sia della specificità del materiale trattato (come nel caso in cui si intenda utilizzare lo standard Dublin Core a un sistema complesso di gestione di documenti archivistici)
- predisporre schemi comprensibili agli uomini e alle macchine (consigliando in questo ultimo caso l'uso di XML)
- documentare gli schemi di metadati sviluppati mediante un sistema di controllo delle versioni che identifichino univocamente ciascuno schema e le sue trasformazioni nel tempo, la registrazione degli schemi all'interno di comunità professionali o di settore, procedure definite di archiviazione che possono prevedere un'architettura centralizzata o distribuita, procedure di monitoraggio e mantenimento dei collegamenti interni agli schemi di metadati o esterni in quanto riferiti ad altre fonti
- stabilire regole per la identificazione logica o l'esportazione fisica dei metadati associati ai documenti
- definire strategie di memorizzazione dei metadati coerenti con le strategie adottate per conservazione permanente dei documenti medesimi.

LO STANDARD ISO 23081.

LE PROCEDURE PER LA CONSERVAZIONE

- Le procedure che sostengono le strategie di conservazione prevedono di:
 - incorporare nei documenti i metadati medesimi (in formati standard quali XML) nei diversi momenti della gestione (al momento della formazione, nei processi di selezione a fini conservativi, nei cambiamenti di supporti di memorizzazione o di ordinamento, nelle fasi di aggiornamento dei sistemi, nei processi di interscambio con altri sistemi, nelle fasi di trasferimento in altri depositi)
 - memorizzare documenti e metadati in formati specifici adatti alla conservazione
 - nel caso di sistemi basati sulla migrazione definire il processo di memorizzazione dei metadati all'interno del software di gestione nei diversi momenti della gestione medesima (al momento della formazione, nei processi di selezione a fini conservativi, nei cambiamenti di supporti di memorizzazione o di ordinamento, nelle fasi di aggiornamento dei sistemi, nei processi di interscambio con altri sistemi, nelle fasi di trasferimento in altri depositi)
 - adottare procedure di conservazione nel tempo dei metadati con particolare attenzione alle esigenze di autenticità, affidabilità e utilizzo e assicurando che ogni passaggio di custodia venga documentato in quanto storia della proprietà dei metadati.